Qiang Xu

+1 (401) 396-6864 | xu1201@purdue.edu | qiangxu1996.github.io | West Lafayette, IN 47907

My research spans across systems, networking, and machine learning. I am interested in **building systems that unlocks the full potential of the underlying hardware resources**. My recent research focuses on building efficient **machine learning inference systems** utilizing GPU resources, both general-purpose and for emerging applications like video analytics and augmented reality.

Education

Purdue University — *West Lafayette, USA*
Ph.D. in Electrical and Computer Engineering — *2018 – 2024 (expected)*
Advisor: Prof. Y. Charlie Hu

University of Science and Technology of China (USTC) — *Hefei, China*
B.E. in Computer Science and Technology — *2014 – 2018*
Advisor: Prof. Yu Zhang
Talent Program in Computer Science and Technology, School of the Gifted Young

Professional Experience

Purdue University — *West Lafayette, USA*
Graduate Research Assistant — *Aug. 2018 – Present*
Advisor: Prof. Y. Charlie Hu

- Designed a machine learning inference framework for heterogeneous GPU clusters. The framework efficiently utilizes old and slower GPUs, improving overall inference throughput by up to 53.2%. (In preprint)
- Developed a machine-learning-as-a-service framework serving augmented reality clients. The framework maximizes the capacity of a GPU server and supports 1.7x–6.9x more clients concurrently. (In preprint)
- Built the first scheduling framework for augmented reality clients that require offloading multiple machine learning tasks. Improved the overall accuracy by 7.6%–14.3%. (**MobiCom 2023**)
- Characterized the performance of offloading object detection tasks over 5G mmWave in the wild in collaboration with wireless networking research teams. (**MASCOTS 2023**, **5G-MeMU 2022**)
- Surveyed 25 mobile app developers for their practices on deep parameters and energy optimization. Systematically studied and categorized the energy impact of deep parameters in 16 Android apps. (**SANER 2022**)
- Contributed to the design of an energy-aware adaptive bitrate algorithm for video streaming. (**USENIX ATC 2021**)

NEC Laboratories America, Inc. — *Princeton, USA*
Research Intern — *May 2023 – Aug. 2023*
Mentor: Murugan Sankaradas

- Designed an offloading scheduler to coordinate DNN-powered video analytics clients under network contention. Reduced the request drop rate by up to 92.9% and improved application responsiveness.

Publications

1. **AROSE: An Accuracy-Aware Proactive Framework for Serving Concurrent Edge-Assisted AR Clients**
   Jonny Kong*, **Qiang Xu***, and Y. Charlie Hu (* co-primary)
   Under submission

2. **Can 5G mmWave Enable Edge-Assisted Real-Time Object Detection for Augmented Reality?**
   Moinak Ghoshal*, Jonny Kong*, **Qiang Xu***, Zixiao Lu, Shivang Aggarwal, Imran Khan, Jiayi Meng, Yuanjie Li, Y. Charlie Hu, and Dimitrios Koutsonikolas (* co-primary)
   31st International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (**MASCOTS 2023**)

3. **AccuMO: Accuracy-Centric Multitask Offloading in Edge-Assisted Mobile Augmented Reality**
   Jonny Kong*, **Qiang Xu***, and Y. Charlie Hu (* co-primary)
   The 29th Annual International Conference on Mobile Computing and Networking (**MobiCom 2023**)

4. **An In-Depth Study of Uplink Performance of 5G mmWave Networks**
   Moinak Ghoshal*, Jonny Kong*, **Qiang Xu***, Zixiao Lu, Shivang Aggarwal, Imran Khan, Yuanjie Li, Y. Charlie Hu, and Dimitrios Koutsonikolas (* co-primary)
   The 2nd ACM SIGCOMM Workshop on 5G and Beyond Network Measurements, Modeling, and Use Case (**5G-MeMU 2022**)

5. **Can 5G mmWave Support Multi-user AR Apps?**
   Moinak Ghoshal, Pranab Dash, Jonny Kong, **Qiang Xu**, Y. Charlie Hu, Dimitrios Koutsonikolas, and Yuanjie Li
   Passive and Active Measurement Conference 2022 (**PAM 2022**)

6. **An Empirical Study on the Impact of Deep Parameters on Mobile App Energy Usage**
   Qiang Xu, James C. Davis, Y. Charlie Hu, and Abhilash Jindal
   The 29th IEEE International Conference on Software Analysis, Evolution and Reengineering (**SANER 2022**)

7. **Do Larger (More Accurate) Deep Neural Network Models Help in Edge-assisted Augmented Reality?**
   Jiayi Meng, Jonny Kong, **Qiang Xu**, and Y. Charlie Hu
   ACM SIGCOMM 2021 Workshop on Network-Application Integration (**NAI 2021**)

8. **Proactive Energy-Aware Adaptive Video Streaming on Mobile Devices**
   Jiayi Meng, **Qiang Xu**, and Y. Charlie Hu
   2021 USENIX Annual Technical Conference (**USENIX ATC 2021**)

Honors and Awards

Ross Fellowship, Purdue University *2018*
National Scholarship (top 0.2% nationwide), USTC *2016*
Outstanding Student Scholarship, USTC *2015*
Outstanding Freshman Scholarship, USTC *2014*

Teaching

Teaching Assistant, Introduction to Operating Systems (ECE 695), Purdue University *2020 – 2023*

Professional Skills

Programming Python, C/C++, Java, JavaScript, MATLAB, Julia, SQL, C#
Platforms Linux, Android, CUDA (TensorRT, Nsight), Docker, HPC
Frameworks PyTorch, Mobile DL frameworks (ncnn, TensorFlow Lite), RL frameworks (Ray, Gym), Unity